

University of Alabama
 University of Alaska, Fairbanks
 University of Arizona
 Arizona State University
 University of Arkansas at Little Rock
 Auburn University
 Baylor University
 Boise State University
 Boston College
 Boston University
 Brown University
 California Institute of Technology
 California State University, East Bay
 California State Polytechnic Univ, Pomona
 University of California, Berkeley
 University of California, Davis
 University of California, Los Angeles
 University of California, Riverside
 University of California, San Diego
 University of California, Santa Barbara
 University of California, Santa Cruz
 Carnegie Institution of Washington
 Central Washington University
 University of Colorado, Boulder
 Colorado School of Mines
 Columbia University
 University of Connecticut
 Cornell University
 University of Delaware
 Duke University
 Florida International University
 University of Florida
 University of Georgia
 Georgia Institute of Technology
 Harvard University
 University of Hawaii at Manoa
 University of Houston
 Idaho State University
 IGGP/Lawrence Livermore Nat. Laboratory
 IGGP/Los Alamos Nat. Laboratory
 University of Illinois at Urbana Champaign
 Indiana University
 Indiana Univ/Purdue Univ, Fort Wayne
 James Madison University
 Kansas State University
 University of Kansas
 University of Kentucky
 Lamar University
 Lawrence Berkeley National Laboratory
 Lehigh University
 Louisiana State University
 Macalester College
 University of Maryland
 Massachusetts Institute of Technology
 University of Memphis
 University of Miami
 Miami University, Ohio
 University of Michigan
 Michigan State University
 Michigan Technological University
 University of Minnesota
 University of Missouri at Columbia
 University of Missouri at Rolla
 Montana Tech of the Univ. of Montana
 University of Nevada, Las Vegas
 University of Nevada, Reno
 New Mexico Inst. of Mining & Technology
 New Mexico State University
 University of New Orleans
 State Univ. of New York at Binghamton
 State Univ. of New York at Stony Brook
 North Carolina State University
 Univ. of North Carolina, Chapel Hill
 Northern Illinois University
 Northwestern University
 Oklahoma State University
 The University of Oklahoma
 University of Oregon
 Oregon State University
 Pennsylvania State University
 University of Puerto Rico, Mayaguez
 Princeton University
 Purdue University
 Rensselaer Polytechnic Institute
 Rice University
 University of Rochester
 Rutgers, State Univ. of New Jersey
 Saint Louis University
 San Diego State University
 San Jose State University
 University of South Carolina
 University of Southern California
 Southern Methodist University
 Stanford University
 Syracuse University
 University of Tennessee
 Texas A&M University
 University of Texas at Arlington
 University of Texas at Austin
 University of Texas at Dallas
 University of Texas at El Paso
 Texas Tech University
 University of Tulsa
 University of Utah
 Virginia Polytechnic Institute
 University of Washington
 Washington University, St. Louis
 West Virginia University
 University of Wisconsin, Madison
 University of Wisconsin, Milwaukee
 University of Wisconsin, Oshkosh
 Western Washington University
 Woods Hole Oceanographic Institution
 Wright State University
 University of Wyoming
 Yale University



November 14, 2014

To: Wendy Wigen (wigen@nitrd.gov)

National Coordination Office for Networking and Information Technology
 Research and Development National Science Foundation

Thank you for the opportunity to provide comments on "The National Big Data R&D Initiative: Vision and Priority Actions". Having reviewed the initial framework, we have responded in the format requested below:

Respondent: Robert Detrick, President, Incorporated Research Institutions for Seismology

Contact: 1200 New York Ave. NW, suite 400, Washington, DC; detrick@iris.edu; 202 682-2220

Experience: IRIS manages data from hundreds of seismographic networks, collectively constituting one of the largest scientific archives of globally distributed observational data in the world. IRIS offers a wide and growing variety of services that Earth scientists rely on in over 150 countries worldwide, increasingly through web services that also deliver a wide range of data products for researchers and public outreach. IRIS manages instrumentation programs that play a key role in facilitating efficient collection of Earth science data and metadata.

What are the gaps that are not addressed in the Visions and Priority Actions document?

The absence of the U. S. Geological Survey or any Department of the Interior agency from the "NITRD Agencies" is a gap that risks lack of attention to a wide variety of *in situ* sensor data for the Earth sciences. The USGS and other DOI agencies work in cooperation with a variety of industries in the commercial sector, and closer cooperation could facilitate development and implementation of new concepts especially well suited to that type of data.

What do you think are the most high impact ideas at the frontiers of big data research and development?

The petroleum industry and other commercial groups have long dealt in very large volumes of Earth science data from ground-based sensor systems, such as 2-D seismic reflection data in the 1980s, and 3-D or 4-D seismic data today. After their proprietary economic value is gone, industry Earth science data are very often lost rather than re-used for public good, partly because of uncertainty about legal ownership. In addition, however, it would be costly to overcome technical hurdles to recover data that were usually not managed for long-term use, and thus were often left in formats and on media that became outdated.

In contrast, Earth science data collected and managed by the U.S. government generally have been managed for ongoing use. Earth science data that have long been important big data include a variety of satellite imagery; gravimetric, synthetic aperture radar, and other geodetic data; and other measurements from remote sensors. Free access to such data has been good for broader society, not only through immediate applications such as weather forecasting and mapping of natural resources but also through fundamental

research using the data. Attempts at cost recovery have failed. For example, when an attempt was made to charge for Landsat data, usage fell so much that costs were not recovered and society lost the benefits of research that was foregone for lack of data access.

Until recently, the volume of seismological and other in situ sensor Earth science data used in academic and government research have grown at a modest pace – for example, much more slowly than the volume of satellite image data. The reason for comparatively slow growth is that academic seismic data have been used as one-dimensional time series, albeit from a growing number of stations. That is changing as dense, unaliased seismic arrays become more important, which is effectively transforming seismic data into three-dimensional data: the time-dependent motion of the Earth's two-dimensional surface.

Wise investment in big seismic data will benefit broader society through reduction of earthquake risk by implementing Earthquake Early Warning Systems and improving earthquake hazard analysis. In addition, seismic data are proving useful in “fluvial seismology” – monitoring river floods and seasonal variations in stream flow – “cryospheric seismology” – monitoring the glacial processes. Seismic monitoring is essential to safe applications of many numerous fluid injection processes, not just to measure earthquakes that might be induced but also through “seismic tomography” to develop numerical descriptions of reservoir behavior through time and test if “sequestered” carbon truly remains buried for significant time periods.

What new infrastructure investments do you think will be game-changing for the big data innovation ecosystem?

The best use of Earth science big data for the benefit of broader society will come, in part, from new investment in *in situ* sensor systems, including seismic systems for academic research that are comparable to systems used today in commercial exploration for petroleum resources. Large numbers of low cost sensors may also be deployed in a citizen-science context, with the data used for hazards analysis. Sensor system investment should be linked to integrated data management, including tools for efficient transfer of data from sensor systems to accessible archives with metadata, use of new storage concepts such as distributed data storage, and high performance compute cycles close to the data, both for data processing and evaluation of numerical simulations.

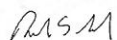
Most current high performance computing facilities have not been designed to truly solve data intensive problems. Rather most current HPC systems are configured with a modeling perspective focusing on large peta or teraflop capability. With the advent of Big Data, new HPC systems, built with high capacity input/output capabilities, need to be made available to the geosciences community. These resources should be made available to the geoscience community either at NSF HPC centers or in collaboration with other government agencies.

How can the federal government most effectively enable new partnerships, particularly those that cross sectors or domains?

Across all aspects of seismological data collection and management, however, there are existing systems already effectively supporting research and societal applications of research-quality data. These systems have been progressively refined over tens of years, and broad research communities have developed modes of working to take advantage of the existing systems. Thus, rather than starting anew, the most effective mode of further investment is to build on the existing systems and facilitate progressive change in work modes among the research user communities.

We are willing to have our ideas posted for discussion on a public Web site and included with possible attribution in the plan.

Sincerely,



Robert Detrick
President